the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2010) **63**, 801–810 doi: 10.1111/j.1365-313X.2010.04284.x

The origin of the non-recombining region of sex chromosomes in *Carica* and *Vasconcellea*

Xia Wu[1], Jianping Wang[2], Jong-Kuk Na[2], Qingyi Yu[3], Richard C. Moore[4], Francis Zee[5], Steven C. Huber[2,6] and Ray Ming[2,*]

[1]*Program in Physiological and Molecular Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 16801, USA,*
[2]*Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA,*
[3]*AgriLife Research Center, Texas A&M University, Weslaco, TX 78596-8344, USA,*
[4]*Department of Botany, Miami University, Oxford, OH 45056, USA,*
[5]*U.S. Department of Agriculture – Agricultural Research Service, Pacific Basin Agricultural Research Center, Hilo, HI 96720, USA, and*
[6]*U.S. Department of Agriculture – Agricultural Research Service, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA*

Received 17 April 2010; revised 7 June 2010; accepted 11 June 2010; published online 19 July 2010.
*For correspondence (fax +1 217 244 1336; e-mail rming@life.illinois.edu).

SUMMARY

***Carica* and *Vasconcellea* are two closely related sister genera in the family Caricaceae, and were once classified as two sections under *Carica*. Sex chromosomes have been found in papaya and originated approximately 2–3 million years ago. The objectives of this study were to determine whether sex chromosomes have evolved in *Vasconcellea*. Six X/Y gene pairs were cloned, sequenced and analyzed from three dioecious, one trioecious and one monoecious species of *Vasconcellea*. The isolation of distinctive X and Y alleles in dioecious and trioecious species of *Vasconcellea* demonstrated that sex chromosomes have evolved in this genus. Phylogenetic analyses indicated a monophyletic relationship between the X/Y alleles of *Carica* and those of *Vasconcellea*. Distinctive clusters of X/Y alleles were documented in *V. parviflora* and *V. pulchra* for all available gene sequences, and in *V. goudatinana* and *V. cardinamarcensis* for some X/Y alleles. The X and Y alleles within each species shared most single nucleotide polymorphism haplotypes that differed from other species. Limited evidence of gene conversion was documented among the X/Y alleles of some species, but was not sufficient to cause the evolutionary patterns reported herein. The *Carica* and *Vasconcellea* sex chromosomes may have originated from the same autosomes bearing the X allelic form that still exist in the monoecious species *V. monoica*, and have evolved independently after the speciation event that separated *Carica* from *Vasconcellea*. Within *Vasconcellea*, sex chromosomes have evolved at the species level, at least for some species.**

Keywords: molecular evolution, papaya, sex chromosomes, single nucleotide polymorphisms haplotype, *Vasconcellea*.

INTRODUCTION

The small family Caricaceae consists of six genera and 35 species, of which 32 are dioecious. The three non-dioecious species are from the genera *Carica* and *Vasconcellea*, and comprise trioecious *C. papaya* and *V. cundinamarcensis* and monoecious *V. monoica*. These two closely related sister genera were once classified as two sections in the genus *Carica*. However, recent molecular and DNA marker data have clearly separated these two sections, resulting in the reinstatement of these two genera (Aradhya *et al.*, 1999; Badillo, 2000; Van Droogenbroeck *et al.*, 2002; Kyndt *et al.*, 2005; Ming *et al.*, 2005). *Carica papaya* is the only species in the genus *Carica*, while *Vasconcellea* is the largest genus in the family and includes 21 species, of which one is monoecious, 19 are dioecious and one is trioecious. These two genera are still considered the most closely related in the family, as inter-generic hybridization is possible through embryo rescue (Manshardt and Wenslaff, 1989). Many traits in *Vasconcellea* have the potential to improve papaya, such as resistance to papaya ringspot virus (PRSV) from *V. caulifora*, the pleasant fragrance of *V. stipulate*, the cold tolerance of *V. cundinamarcensis*, the ornamental qualities of pink-flowered

V. parliflora and the monoecious habit of *V. monoica* (Manshardt and Wenslaff, 1989).

Dioecy results in out-crossing, which enhances genetic diversity and is known to have a fitness advantage. Sex chromosomes have evolved multiple times independently in animals, plants and fungi (Graves and Shetty, 2001; Fraser *et al.*, 2004; Ming and Moore 2007). Sex chromosomes reinforce dioecy and result from selective pressure to link two sex determination mutations, one promoting maleness and the other suppressing femaleness (Charlesworth and Charlesworth, 1978; Charlesworth, 1991). These sex determination loci were permanently linked when suppression of recombination occurred in this region and a proto-Y (or Z) chromosome was formed. The absence of recombination leads to accumulation of deleterious mutations, degeneration of the Y (or Z) chromosome, and the formation of heteromorphic sex chromosome pairs as seen in ancient sex chromosome systems. A broad definition of sex chromosomes could be simply that they contain sex determination genes, even without suppression of recombination at the sex determining region, as in strawberry (Spigler *et al.*, 2008). A more strict definition would include suppression of recombination at the sex-determining region and degeneration of its gene content, as demonstrated in ancient human sex chromosomes and nascent sex chromosomes in plants and fish (Skaletsky *et al.*, 2003; Liu *et al.*, 2004; Peichel *et al.*, 2004; Yin *et al.*, 2008).

Papaya is one of the two trioecious species in the family, and sex determination is controlled by primitive sex chromosomes (Liu *et al.*, 2004). When compared with the 166 million year old human sex chromosomes, papaya sex chromosomes are at the beginning of their evolution, which started approximately 2–3 million years ago (Mya) (Yu *et al.*, 2008a). Dioecy appears to be ancestral in the family Caricaceae, as no hermaphroditic species exist in the family. Papaya sex chromosomes are probably evolved from an ancestral sex determination system that is shared at least with *Vasconcellea*, if not the entire family, because intergeneric crosses between *Carica* and *Vasconcellea* species showed no functional complementation for sex determination, indicating that their sex determination genes are allelic (Magdalita *et al.*, 1997).

Papaya has an XY system with two slightly different Y chromosomes, a male-specific Y and a hermaphrodite-specific Y^h (Ming *et al.*, 2007a). At least one X chromosome is required for the survival genotypes: XX female, XY male, and XY^h hermaphrodite (YY, Y^hY^h and YY^h are embryo-lethal). Moreover, the Y and Y^h chromosomes are mutually exclusive, and no papaya has been found with an XYY, XYY^h or XY^hY^h genotype. Both the Y and Y^h chromosomes contain a small male-specific region (MSY) of approximately 8–9 Mb with low gene density and a high level of repetitive sequences (Yu *et al.*, 2007, 2008a,b; Ming *et al.*, 2008). The MSY in papaya is pericentric and probably includes the centromere (Zhang *et al.*, 2008).

The first six papaya sex-linked genes were identified in two MSY BACs and two corresponding X BACs, including four genes on the MSY BACs and six genes on the X BACs (Yu *et al.*, 2008a). These six genes were named based on their sequential location on the BACs following the example of the *Silene* sex-linked genes (Filatov and Charlesworth, 2002; Nicolas *et al.*, 2005). Four X/Y^h gene pairs were analyzed and the estimated divergence time between the X and Y^h chromosomes was 2–3 million years (Yu *et al.*, 2008a), perhaps after the speciation event of *C. papaya* that occurred more than 3 Mya (Aradhya *et al.*, 1999). These results indicate that sex chromosomes cannot be ancestral in the family Caricaceae. Recombination between the X and Y chromosomes in the MSY region may have been suppressed over the past 2–3 million years, and the X and Y sequences have diverged, sharing approximately 83–86% genomic sequence identity (Yu *et al.*, 2008a,b). However, the Y and Y^h homologous BACs shared 98.8% sequence identity, including the repetitive sequences, and the estimated time of divergence is approximately 73 000 years (Yu *et al.*, 2008b). Given the prevalence of dioecy in the family Caricaceae, hermaphroditic papayas probably evolved from the males with the Y chromosome.

Currently, it is not known whether sex chromosomes exist in other genera of Caricaceae. The only species that is certain to have no sex chromosomes is the monoecious species *V. monoica*, because there is no sexual dimorphism among individual plants. The objective of this study was to assess the MSY homologous regions in selected *Vasconcellea* species to test whether sex chromosomes have evolved in *Vasconcellea*, and if so, whether the sex chromosomes originated before or after the speciation events of these two genera and within *Vasconcellea*.

RESULTS

The six X/Y gene pairs selected for this experiment are located on two pairs of papaya X/Y homologous BACs, genes 1–3 on X/Y BACs SH53E18 and SH85B24 (GenBank accession numbers EF661026 and EF661024) and genes 4–6 on X/Y BACs SH61H02 and SH95B12 (GenBank accession numbers EF661023 and EF661025). These two pairs of BACs are 4–5 Mb apart, and span approximately 60% of the 8–9 Mb papaya MSY, and the gene numbers are based on their positions on the BACs (Figure S1). The X/Y allelic sequences of the six target genes were amplified from *Vasconcellea* species, including genes 1 and 3 that have not been reported before. Distinct X and Y alleles of gene 1 were obtained from three dioecious and one trioecious species, but only the X allelic form was obtained from the monoecious species *V. monoica*. For the other five gene pairs, we only obtained fragments of both sex types from one to three *Vasconcellea* species, despite numerous attempts (Tables

S1 and S2). This difficulty was mainly attributed to large introns and possible sequence variations. The autosomal gene *ANKYRIN* was amplified from all five *Vasconcellea* species with only one allelic form.

The X alleles of genes 1–3 in papaya were located on the X-specific BAC SH53E18 (Table S3), and the Y allele of gene 2 was on the Y-specific BAC SH85B24 (gene 3 was deleted from the MSY) (Yu *et al.*, 2008a). The Y allele of gene 1 was not identified until MSY BAC SH69A08 was sequenced recently (GenBank accession number AM778093). The availability of this sequence enabled us to locate gene 1, which was 1.7 Mb away from BAC SH85B24. The X allele of papaya gene 1 encoded 684 amino acids in 19 exons, and the closest match to the sequence is a protein kinase-like protein from *Arabidopsis thaliana* (AT5G14720). In contrast, the Y allele encoded 678 amino acids, with a deletion of six amino acid residues in exon 15 (truncated sequence LLNFTA). As gene 1 was predicted to be a kinase-like protein, we also performed a sequence search in the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB). The N-terminal amino acid sequence of gene 1 had 43% identity (125 of 292 residues) with human STE20-like kinase 3 (MST3, STK24, PDB accession number 2j51; Pike *et al.*, 2008), which is on the long arm of the human X chromosome (National Center for Biotechnology Information accession HsX_11943).

Partial coding sequences of the six X/Y gene pairs (genes 1–6) and the autosomal gene *ANKYRIN* were analyzed to determine the degree of synonymous (K_s) and non-synonymous (K_a) divergence within and between species. The ratio of non-synonymous to synonymous divergence (K_a/K_s) between the X and Y alleles can be used to infer the degree of functional constraint and the level of divergence. The total number of synonymous and non-synonymous sites, mutations and the degree of divergence are summarized in Table 1 and Tables S4–S8. All six X/Y gene pairs had K_a/K_s ratios <1, indicating functional constraint. For gene 1, which was obtained in all six species, K_a/K_s ratios ranged from 0.07 to 0.51 for comparison of X versus Y alleles within species, from 0 to 0.31 for comparison of X versus X alleles between species, and from 0.07 to 0.74 for comparison of Y versus Y alleles between species.

In addition to the K_a/K_s ratios, the number of synonymous and non-synonymous mutations is also an indication of the level of divergence. Within *Vasconcellea* species, the number of synonymous mutations between X and Y alleles ranged from 0 to 8, and the number of non-synonymous mutations ranged from 1 to 7 (Table 1). Between genera (i.e. for *Carica papaya* versus *Vasconcellea* species), the numbers of synonymous and non-synonymous mutations increased to 13–30 and 6–24, respectively, for the X allele, and to 12–32 and 8–23 for the Y allele. Between *Vasconcellea* species, the numbers of synonymous and non-synonymous mutations were 1–6 and 0–3, respectively, for the X allele, and 0–12 and 0–8 for the Y allele. Similar patterns were observed for the other five X/Y gene pairs (Tables S4–S8).

For the autosomal gene *ANKYRIN*, the K_a/K_s ratios ranged from 0.05 to 0.13 between *Carica* and *Vasconcellea* species and from 0.32 to 0.49 among *Vasconcellea* species (Table 2). The numbers of synonymous and non-synonymous mutations were 12 and 2, respectively, for all five pairwise comparisons between *Carica* and *Vasconcellea* species, and ranged from 0 to 2 and 0 to 3, respectively, among *Vasconcellea* species. It should be noted that six pairwise comparisons of *ANKYRIN* showed no synonymous or non-synonymous mutations among *Vasconcellea* species.

Divergence times of X/Y pairs from each species were determined using the uncorrelated Bayesian relaxed-clock model, with branch rates independently drawn from an exponential distribution with a mean substitution rate of 7.1×10^{-9} substitutions per site per year (Ossowski *et al.*, 2010), as implemented in the software BEAST 1.4 (Drummond *et al.*, 2006; Drummond and Rambaut, 2007). For *C. papaya*, mean divergence times ranged from 0.57 to 5.4 Mya for X/Y alleles (Table 3), consistent with previous reports of the relatively recent divergence of the X and Y chromosome in this species (Yu *et al.*, 2008a,b). The mean divergence time of the *C. papaya* male Y and hermaphroditic Y^h was consistently more recent than the X/Y divergence, ranging from 0.13 to 0.32 Mya, supporting the hypothesis that hermaphrodites derived from males in the recent past. The estimated divergence times of the allele pairs for *Vasconcellea* species was similar to that found for *C. papaya* X/Y alleles. Estimated divergence times ranged from 0.89 to 4.4 Mya for *V. goudotiana*, 0.80 to 3.1 Mya for *V. pulchra*, 0.7 to 2.0 Mya for *V. parviflora* and 0.64 to 2.8 Mya for *V. cundinamarcensis*. For all genes, the divergence times for all X/Y pairs were more recent than the mean divergence time of *Carica* from *Vasconcellea* species (6.8–14.6 Mya), indicating independent evolution of X/Y alleles in these genera.

We tested the possibility of evolutionary strata on the *C. papaya* MSY by comparing the mean estimated divergence times of the X/Y alleles of genes 1–3 with the mean estimated divergence of genes 4–6; these groups of genes are found on separate BAC clones and are physically separated on the MSY by 4–5 Mb. Although the mean divergence time for genes 4–6 (3.2 ± 1.1 Mya) was almost twice that of genes 1–3 (1.7 ± 0.30 Mya), these ages were not significantly different (*t* test, $P > 0.05$). Thus, the presence of evolutionary strata is not supported based on this limited survey of MSY genes.

We analyzed the phylogenies of X/Y alleles (or triplets for papaya) among *C. papaya* and *Vasconcellea* species using Bayesian, maximum-likelihood, maximum-parsimony and neighbor-joining methods (Figure 1 and Figures S2–S4). Except for gene 4 (exons 1 and 2), *C. papaya* X, Y and Y^h alleles formed a cluster that was clearly separated from their

Table 1 Estimates of synonymous and non-synonymous nucleotide divergence within and between species for gene 1

Species	Total sites	Total coding sites	Synonymous sites	Non-synonymous sites	Synonymous mutations	Non-synonymous mutations	K_s	K_a	K_a/K_s
Within species, X versus Y alleles									
C. papaya (*Cp*)	10 732	2034	457.3	1576.8	4	7	0.009	0.004	0.51
V. cundinamarcensis (*Vc*)	766	255	59.8	195.2	0	2	0.000	0.010	–
V. goudotiana (*Vg*)	2076	768	179.9	588.1	8	2	0.046	0.003	0.07
V. parviflora (*Vpa*)	1680	510	123.8	386.2	3	3	0.025	0.008	0.32
V. pulchra (*Vpu*)	782	273	64.4	208.6	2	1	0.032	0.005	0.15
Between species, X allele									
Cp versus *Vc*	885	375	91.3	283.8	16	6	0.200	0.021	0.11
Cp versus *Vg*	2951	990	233.2	756.8	30	24	0.141	0.031	0.22
Cp versus *Vm*	2104	795	185.1	609.9	26	10	0.156	0.017	0.11
Cp versus *Vpa*	1739	576	140.3	435.8	16	19	0.124	0.045	0.36
Cp versus *Vpu*	826	321	75.1	245.9	13	6	0.197	0.025	0.13
Vc versus *Vg*	886	375	92.2	282.8	1	0	0.011	0.000	0.00
Vc versus *Vm*	888	375	92.3	282.7	2	1	0.022	0.004	0.16
Vc versus *Vpa*	887	372	91.5	280.5	4	0	0.045	0.000	0.00
Vc versus *Vpu*	826	321	75.9	245.1	5	1	0.069	0.004	0.06
Vg versus *Vm*	2116	783	184.8	598.2	3	3	0.016	0.005	0.31
Vg versus *Vpa*	910	396	98.2	297.8	3	0	0.031	0.000	0.00
Vg versus *Vpu*	825	321	75.9	245.1	5	1	0.069	0.004	0.06
Vm versus *Vpa*	1761	519	125.9	393.1	5	3	0.041	0.008	0.19
Vm versus *Vpu*	827	321	76.1	244.9	6	2	0.083	0.008	0.10
Vpa versus *Vpu*	828	321	75.9	245.1	4	1	0.055	0.004	0.07
Between species, Y allele									
Cp versus *Vc*	762	255	58.7	196.3	12	8	0.239	0.042	0.18
Cp versus *Vg*	3046	999	233.8	765.2	32	23	0.151	0.033	0.22
Cp versus *Vm*	2067	768	178.1	589.9	28	10	0.176	0.017	0.10
Cp versus *Vpa*	1610	456	107.2	348.8	15	15	0.155	0.044	0.29
Cp versus *Vpu*	779	279	63.8	215.2	12	6	0.216	0.028	0.13
Vc versus *Vg*	790	255	59.8	195.3	5	3	0.089	0.016	0.17
Vc versus *Vm*	767	252	59.7	192.3	0	3	0.000	0.016	–
Vc versus *Vpa*	768	252	59.5	192.5	4	2	0.070	0.010	0.15
Vc versus *Vpu*	760	252	59.8	192.2	3	2	0.052	0.010	0.20
Vg versus *Vm*	3027	999	238.5	760.5	12	8	0.052	0.011	0.20
Vg versus *Vpa*	1138	438	105.4	332.6	8	4	0.080	0.012	0.15
Vg versus *Vpu*	782	294	69.3	224.8	4	1	0.060	0.004	0.07
Vm versus *Vpa*	1739	510	123.3	386.8	8	7	0.068	0.018	0.27
Vm versus *Vpu*	781	273	64.5	208.5	4	1	0.065	0.005	0.07
Vpa versus *Vpu*	782	273	64.3	208.7	3	0	0.048	0.036	0.74

Vm, *V. monoica.*

Vasconcellea homologs in five of the six phylogenetic trees. Among *C. papaya* X, Y and Y^h alleles, the Y and Y^h alleles formed a sub-cluster before joining the X allele in genes 4 (exons 7–18), 5 and 6.

The X/Y alleles for genes 1, 2 and 4 formed a distinct cluster in *V. pulchra*, as did the X/Y alleles for genes 1, 3 and 4 in *V. parviflora*. The X/Y alleles of genes 1, 2, 4 and 6 were amplified from *V. goudotiana*, and genes 2 and 4 formed a distinct cluster, but genes 1 and 6 did not. Similarly, the X/Y alleles of genes 1 and 6 were amplified from *V. cundinamarcensis*: gene 6 formed a distinct cluster, but gene 1 did not.

To verify the phylogenetic relationships, we examined single nucleotide polymorphisms (SNPs) between *Carica* and *Vasconcellea* X and Y alleles. If the sex chromosomes evolved after a speciation event, the SNP haplotypes would differ between species, but not within species. Indeed, this type of SNP haplotype is prevalent in our data (Figure S5). For example, at gene 1, exon 1, bp 45, papaya X and Y have GAA, encoding glutamic acid, whereas all *Vasconcellea* species have GAT, encoding aspartic acid. At gene 1, exon 1, bp 51, both papaya X and Y have AAA, coding for lysine, but all the *Vasconcellea* species share AAG, which also encodes lysine. Twenty polymorphic synonymous mutations and three non-synonymous mutations were found with this pattern in the 800 bp of exons 1 and 2 of gene 1. For the exon fragments of other genes, the SNP haplotypes of papaya were also more identical between X and Y alleles than across species. Only a few exceptions

Table 2 Estimates of synonymous and non-synonymous nucleotide divergence within and between species for the autosomal *ANKYRIN* gene

Species	Total sites	Total coding sites	Synonymous sites	Non-synonymous sites	Synonymous mutations	Non-synonymous mutations	K_s	K_a	K_a/K_s
Between species									
Cp versus *Vc*	1477	453	111.2	341.8	12	2	0.117	0.006	0.05
Cp versus *Vg*	1583	456	111.8	344.2	12	2	0.116	0.006	0.05
Cp versus *Vm**	1582	456	111.8	344.2	12	2	0.116	0.006	0.05
Cp versus *Vpa*	1579	456	111.8	344.2	12	2	0.116	0.006	0.05
Cp versus *Vpu*	1581	456	112.0	344.0	12	5	0.116	0.015	0.13
Vc versus *Vg*	1549	453	111.3	341.7	0	0	0.000	0.000	–
Vc versus *Vm*	1550	453	111.3	341.7	0	0	0.000	0.000	–
Vc versus *Vpa*	1547	453	111.3	341.7	0	0	0.000	0.000	–
Vc versus *Vpu*	1545	453	111.5	341.5	2	3	0.018	0.009	0.49
Vg versus *Vm*	1660	456	112.0	344.0	0	0	0.000	0.000	–
Vg versus *Vpa*	1653	456	112.0	344.0	0	0	0.000	0.000	–
Vg versus *Vpu*	1650	456	112.2	343.8	2	3	0.018	0.009	0.49
Vm versus *Vpa*	1653	456	112.0	344.0	0	0	0.000	0.000	–
Vm versus *Vpu*	1650	456	112.2	343.8	2	2	0.018	0.006	0.32
Vpa versus *Vpu*	1648	456	112.2	343.8	2	3	0.018	0.009	0.49

Cp, *C. papaya*; Vc, *V. cundinamarcensis*; Vg, *V. goudotiana*; Vpa, *V. parviflora*; Vpu, *V. pulchra.*

Table 3 Estimated divergence times (Mya) for X/Y alleles in *C. papaya* and *Vasconcellea* species

	Gene ID						
Comparison[a]	1	2	3	4 (region 1)	4 (region 2)	5	6
C. papaya	1.4[b] (0.23, 3.4)[c]	2.0 (0.36, 4.5)	–	5.4 (1.2, 12.6)	4.5 (1.0, 10.4)	2.5 (0.52, 6.1)	0.57 (0.03, 1.5)
C. papaya (Y/Y^h)[d]	ND	0.19 (0, 0.64)	–	–	–	0.13 (0, 0.47)	0.32 (0, 1.0)
V. goudotiana	4.4 (1.4, 9.6)	0.89 (0.09, 2.1)	–	–	0.74 (0.04, 2.2)	–	3.2 (0.79, 7.2)
V. pulchra	0.80 (0.15, 1.8)	0.44 (0.005, 1.3)	–	3.1 (0.56, 8.0)	–	–	–
V. parviflora	1.0 (0.21, 2.3)	–	2.0 (0.21, 5.4)	–	–	0.70 (0.06, 1.8)	–
V. cundinamarcensis	2.8 (0.66, 5.8)	–	–	–	–	–	0.64 (0.09, 1.5)
C. /Vasoncellea sp.[e]	13.9 (4.4, 28.7)	13.2 (4.7, 26.4)	10.5 (3.0, 22.3)	14.7 (4.4, 31.6)	13.2 (4.3, 27.4)	7.2 (2.3, 14.5)	6.8 (2.4, 14.1)

[a]Comparisons between X and Y alleles unless otherwise noted.
[b]Mean estimated divergence time in millions of years.
[c]Lower and upper bounds of the 95% highest posterior density interval.
[d]Divergence time estimates for the male Y and hermaphroditic Y^h.
[e]Estimated mean divergence time of *C. papaya* and *Vasconcellea* species.

were found in which only one allelic form of papaya was mutated while the other remained identical to those of *Vasconcellea* species. At gene 1, exon 1, bp 30, the papaya Y allele has a CCT synonymous mutation encoding proline, while the *Vasconcellea* species have a CCC codon, the same as papaya X allele. Likewise, at gene 1, exon 1, bp 105, where the papaya X allele was TGT (cysteine), but the Y allele was TGC (cysteine), which is the codon found in all five *Vasconcellea* species. Nevertheless, these exceptions do not change the general conclusion that the $X/Y(Y^h)$ chromosome system in papaya evolved independently of the *Vasconcellea* genus.

SNP haplotypes from *V. pulchra* and *V. parviflora* appeared to support independent evolution of nascent sex chromosomes in each of these two species. At gene 2, exon 6, bp 141, both the X and Y alleles of *V. pulchra* had codon CTA, coding for leucine, whereas the other four *Vasconcellea* species had the synonymous codon CTC. At gene 5, exon 11, bp 323, the X and Y alleles of *V. parviflora* had codon GTA, coding for proline, whereas the other four *Vasconcellea* species had the synonymous codon GTG.

Close examination of gene 1 sequences showed that most variations were at the C-terminal exons. For the allele of the monoecious species *V. monoica*, a deletion of two amino acid residues (KY) occurred at bp 23–28 of exon 16. For both the X and Y alleles of *V. parviflora*, an insertion of guanine at bp 12 of exon 18 resulted in a frameshift mutation such that the sequence PSTPRERELMAQ was amended to RFYSWRERAHSA. For the X allele of *V. goudotiana*, an SNP at bp 44 of exon 15 resulted in a change from TCA (Y allele sequence)

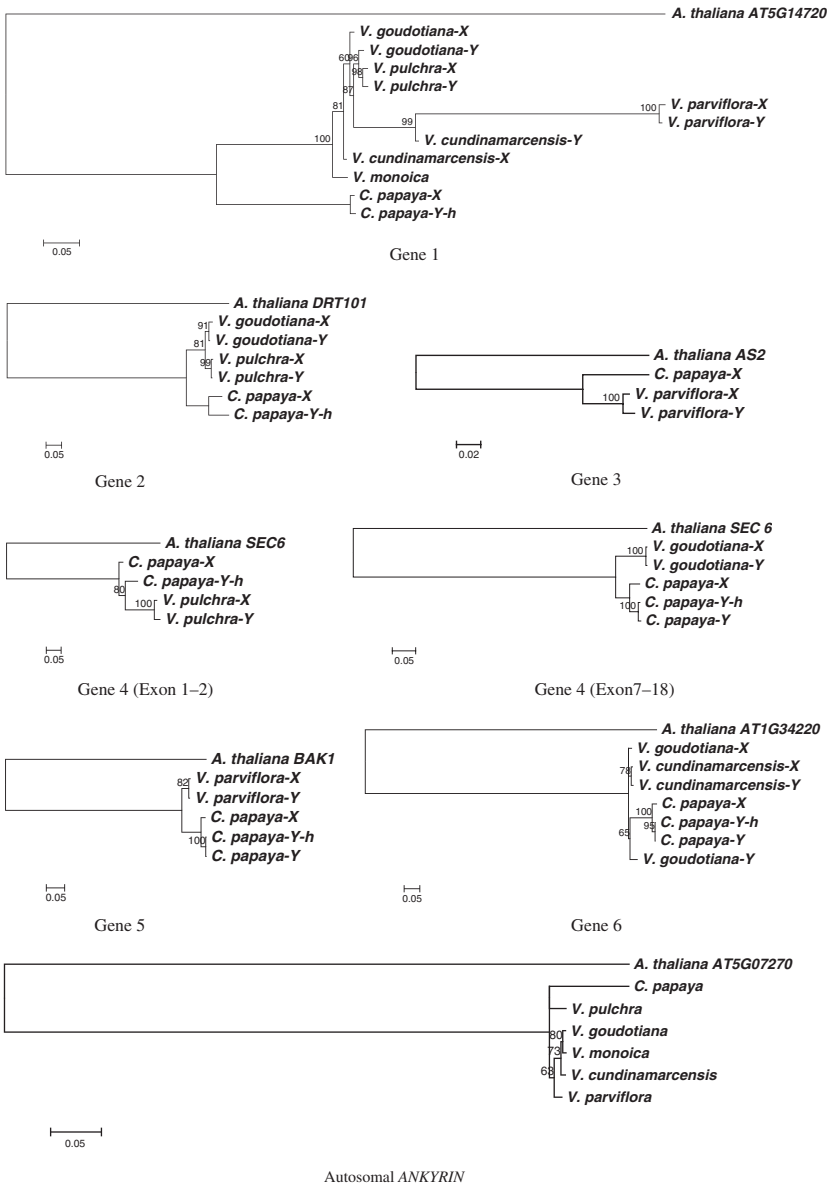

Figure 1. Bayesian phylogenetic trees of papaya sex-linked genes 1–6 and the autosomal gene *ANKYRIN*.
All trees were generated from genomic sequence alignments (using all sites except gaps) using the MrBayes (3.1.2) method (Huelsenbeck and Ronquist, 2001). Other methods (maximum-likelihood, maximum-parsimony and neighbor-joining) yielded similar results. Branch lengths are shown in proportion to total sequence divergence. Numbers at nodes are bootstrap values that exceed 50% (based on 1000 replicates). Corresponding homologs of *A. thaliana* were used as the outgroup. X and Y represent the X- and Y-specific alleles, respectively. For papaya, Y represents the male Y allele and Y-h represents the hermaphrodite Y^h allele.

encoding serine, to TAA (X allele sequence), which is a premature stop codon.

We examined the possibility of gene conversion of these X/Y gene pairs using GENECONV software (Sawyer, 1989, 1999). No gene conversion was detected in gene 3 or the autosomal gene *ANKYRIN*. For the remaining five genes, only short fragments in six of the 24 pairs of sequences within species showed gene conversion, including 707 bp of gene 1 in *V. parviflora*, 111 bp of gene 2 in *V. pulchra*, 95 bp of gene 4 in papaya, 252 bp of gene 5 in papaya, 157 bp of of gene 6 in *V. cundinamarcensis*, and 371 bp of gene 6 in *V. goudotiana* (Table S9). We reconstructed the phylogenetic trees after removing the sequences with gene conversion, and the resulting trees were the same as in Figure 1 (Figure S6). The limited amount of gene conversion corroborated our conclusion that the divergence between MSY and its X counterpart is the consequence of recombination suppression at the sex-determining region over the course of sex chromosome evolution.

A summary diagram representing the evolution of the sex chromosomes in the *Vasconcellea* and *C. papaya* clade in a simplified form is shown in Figure 2.

DISCUSSION

Sexual reproduction increases genetic diversity and generates novel allele combinations for natural selection to act on (Weismann, 1889; Goddard *et al.*, 2005; Paland and Lynch, 2006). The emergence of sex chromosomes requires the suppression of recombination in the genomic region harboring the sex determination genes, which leads to degeneration of the Y (if male heterogametic) or Z (if female heterogametic) chromosome and the divergence of X/Y (or

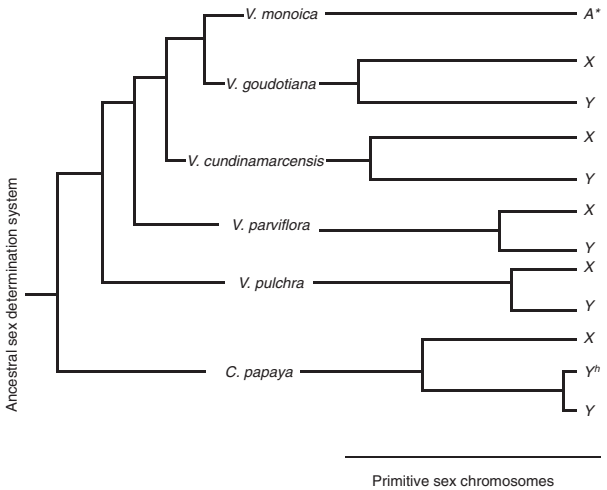

Figure 2. Composite summary diagram based on Figure 1 showing evolution of the sex chromosomes in *Vasconcellea* and *Carica* clade in a simplified form. A*, autosomal allele.

Z/W) alleles (Charlesworth and Charlesworth, 1978; Charlesworth, 1991; Ming *et al.*, 2007b). After the primitive sex chromosomes were uncovered in papaya (Liu *et al.*, 2004), the question arose as to whether such sex chromosomes exist in other genera of Caricaceae. We investigated selected species of *Vasconcellea*, the closest relatives of *Carica*, using six sex-linked genes cloned in papaya (Yu *et al.*, 2008a,b). Distinctive X- and Y-specific alleles in three dioecious and one trioecious *Vasconcellea* species were identified, whereas only one allelic form was identified in the monoecious species *V. monoica*. Synonymous and non-synonymous mutations were found between X and Y alleles, and the estimated divergence times between the X and Y alleles in *Vasconcellea* species ranged from 0.64 to 4.4 Mya. In contrast, only one allelic form of the autosomal gene *ANKYRIN* was found in each species, and no synonymous and non-synonymous mutations were found in six of the ten pairwise comparisons among five *Vasconcellea* species. Our results demonstrated that sex chromosomes have evolved in at least some, if not all, dioecious and trioecious *Vasconcellea* species.

The six X/Y gene pairs used in this experiment are 4–5 Mb apart, spanning more than half the 8–9 Mb papaya MSY (Yu *et al.*, 2008a). The identification of X and Y allelic forms of six orthologous genes in *Carica* and *Vasconcellea* indicated that their sex chromosomes evolved from the same ancestral autosomes. It is likely that *Carica* and *Vasconcellea*, if not the entire family, share the same ancestral sex determination genes that triggered sex chromosome evolution.

Our next question was whether sex chromosomes evolved before or after the speciation event of the two genera. If it occurred before the speciation event, all alleles of the same sex type regardless of the species should be clustered together; if it happened after the speciation event, the X and Y alleles within species would be clustered together but separated from the X and Y allele clusters of other species, as shown for the most recently diverged X/Y gene pairs on the sex chromosomes of selected *Silene* species (Nicolas *et al.*, 2005). Our results showed that the papaya X, Y and Y^h alleles form a distinct cluster from other X and Y alleles of *Vasconcellea* species (Figure 1), showing that the sex chromosomes in *Carica* evolved after the divergence of these two genera. Furthermore, the estimated divergence time of X/Y alleles in *C. papaya* was more recent than the divergence time of these two genera (Table 3), demonstrating that X/Y alleles diverged after the papaya speciation event. Moreover, the *Carica* X and Y alleles have SNP haplotypes that are different from those of *Vasconcellea* X and Y alleles in all but a few cases. The phylogenetic analysis, divergence estimation and SNP analysis all strongly indicated that the sex chromosomes in *Carica* and *Vasconcellea* have evolved independently after the speciation event that separated these two genera. This is in line with the notion that the ancestor of *C. papaya* moved from South America to Central America across island chains before formation of the Isthmus of Panama 3 million years ago (Aradhya *et al.*, 1999). The ancestor of papaya continued to evolve in isolation from its sister genus *Vasconcellea* into a distinct species before the sex chromosomes emerged.

The estimated divergence time between the X and Y alleles in *Carica* is on average higher than that within *Vasconcellea* species, although this difference is not significant (ANOVA, $P > 0.05$). This is in line with the fact that *Carica* and *Vasconcellea* evolved into separate genera before the speciation events within *Vasconcellea*. The number of synonymous and non-synonymous mutations between X and Y alleles in papaya is consistently higher than in *Vasconcellea* species. Furthermore, the Y allele of gene 3 appeared to be degenerated in *Carica*, but still exists in *Vasconcellea*. These data confirm that the *Vasconcellea* sex chromosomes in some species may have evolved more recently than in *C. papaya*, particularly the chromosomes of *V. pulchra*, *V. parviflora* and *V. cundinamarcensis*, with mean divergence times between 1.2 and 2.3 Mya, compared with the mean divergence of 2.7 Mya in *C. papaya* (Table 3).

We also addressed the alternative hypothesis that gene conversion has homogenized X/Y allele sequences and is the cause of the lower divergence of X/Y alleles within *Carica* and *Vasconcellea* species than between species. Although we did find evidence for tracts of conversion in some X/Y allele pairs, our analysis does not support this alternative hypothesis (Table S9). First, the detected conversion events were inconsistently distributed among X/Y allele pairs of each species for any gene analyzed. For example, in gene 1, conversion was detected among the X/Y alleles of *V. pavifl-ora*, but not in *C. papaya* or other *Vasconcellea* species. Similarly, in other genes, we found evidence for conversion

in one or two species, but not all species analyzed. Second, the detected gene conversion events were limited to tracts of a hundred to a few hundred base pairs, and none spanned the entire length of the gene. Importantly, we obtained the same evolutionary relationships for X/Y alleles when we removed the conversion tracts from our alignments. Thus, while we did find evidence of limited gene conversion among X/Y alleles for some species, we suggest that the conversion was not sufficient to cause the evolutionary patterns seen among X/Y alleles. Instead, we suggest that these patterns are caused by independent origins of suppressed recombination in the genetic regions containing these genes in the various species.

Papaya is the only species in the *Carica* genus, but there are 21 species in *Vasconcellea*, including monoecious, dioecious and trioecious species (Ming *et al.*, 2005). The fact that a monoecious species exists in this genus suggests that sex chromosomes are not ancestral for the entire genus, although reverse evolution could account for the monoecious species. Our data support the conclusion that sex chromosomes evolved independently at the species level in *V. parviflora* and *V. pulchra*, and probably in *V. goudotiana* and *V. cundinamarcensis* as well, because distinctive clusters of X/Y alleles formed in all genes analyzed in *V. parviflora* and *V. pulchra* and some genes in *V. goudotiana* and *V. cundinamarcensis*.

Is it possible that these six genes are located in a recently evolved evolutionary stratum but the sex chromosomes are ancient? Genes 1–3 and genes 4–6 are 4–5 Mb apart across more than half the MSY, but the mean estimated divergence times of X/Y alleles in *C. papaya* are not significantly different. We cannot test for strata in *Vasconcellea* sp., however, as details of the physical and/or genetic positions of these loci are lacking in these species. We hypothesize that the MSY region in *Carica* and *Vasconcellea* probably expanded from initial recombination suppression in a region containing a shared pair of sex determination genes, but appears to have occurrred too recently for formation of detectable evolutionary strata.

Although sex chromosomes in these two genera evolved recently, degeneration of the Y chromosome is already in progress. In addition to the Y allele of gene 3 being deleted in papaya, the Y allele of gene 1 in *V. goudotiana* may be a pseudogene due to a point mutation resulting in a premature stop codon.

EXPERIMENTAL PROCEDURES

Plant materials and sex allele identification

Five *Vasconcellea* species were selected for this experiment, including one trioecious species, *V. cundinamacensis*, three dioecious species, *V. parviflora*, *V. goudotiana* and *V. pulchra*, and one monoecious species, *V. monoica*. Genomic DNA was extracted using standard protocols from different sex types of the five *Vasconcellea* species, together with the gynodioecious papaya variety 'SunUp' and the dioecious papaya 'Au9'. For allele identification in *Vasconcellea* species, the X allele sequences were obtained from female plants with the XX genotype; the Y allele sequences were obtained from male plants with the XY genotype if clear sequence variations were observed after comparing with the X allele of the same species. For papaya, sex-specific alleles were obtained from the sex-specific BACs (Yu *et al.*, 2008a).

Degenerate primer design and degenerate PCR

Six papaya sex-linked genes were selected for amplifying homologous regions of *Vasconcellea* species (Yu *et al.*, 2008a). The autosomal *ANKYRIN* gene was used as a control. The conserved regions of each gene were used to design degenerate primers with the criteria of at least 20 bp length and <96 degeneracy. Inosines were used as a neutral base match at the most divergent positions. Partial codons were included at the end of the primers for specific matches (Table S1).

The degenerate primers were used to amplify target genes from different sex types of each selected *Vasconcellea* species. The degenerate PCR reaction include 1 × reaction buffer, 150 μM dNTP, 1.25 μM degenerate primer or 0.5 μM specific primer, 2 units EconoTaq (#30031-1; Lucigen) and 5 ng genomic DNA template. Amplification was performed using the program: 94°C for 5 min; 40 cycles of 94°C for 30 sec, 50°C for 45 sec and 72°C for between 40 sec and 2 min (according to the PCR product size), and 72°C for 10 min.

Cloning and sequencing

The specific PCR products matching the target size were purified using a PCR Clean-Up Kit (#A9282; Promega, http://www.promega.com/) and cloned into pGEM-T EASY vector (#A1360; Promega) according to the manufacturer's instructions. The ligation products were transformed into JM109-competent cells (Promega). Positive colonies were confirmed by PCR using T7-pro primer (5′-GTAATACGACTCACTATAGGGC-3′) and M13Rev-48 primer (5′ GAACAGCTATGACCATGATTAC-3′). Plasmids of positive colonies were isolated and adjusted to a concentration of 100–200 ng μl^{-1} for sequencing using the T7-pro and M13Rev-48 primers at the University of Illinois at Urbana-Champaign Core Sequencing Facility. Each gene fragment was sequenced at least twice to eliminate any sequence error.

Primer walking

Once the gene sequences from *Vasconcellea* species were available, sequence-specific primers (Table S2) were designed based on the conserved regions to amplify the genes in those species that could not be amplified by degenerate primers. Nested sequence-specific primers were designed to further extend the sequences.

DNA sequence divergence analysis

Pairs of sequences were sequentially aligned using the BioEdit program (Hall, 1999). The alignments were exported in.nex format file for analysis using the DnaSP 4.20 program (Rozas *et al.*, 2003). Exons and introns of the X and Y gene pairs amplified from *Vasconcellea* species were designated based on the papaya gene structure. The numbers of synonymous substitutions per synonymous site (K_s), non-synonymous substitutions per non-synonymous site (K_a), and synonymous and non-coding substitutions per silent site (K_{sil}) were calculated as described by Nei and Gojobori (1986) and implemented in DnaSP 4.20 (Rozas *et al.*, 2003).

Divergence times of X/Y pairs from each species were determined using the program BEAST version 1.4 (Drummond *et al.*, 2006; Drummond and Rambaut, 2007). We used an uncorrelated Bayesian relaxed-clock model with branch rates independently drawn from an exponential distribution with a mean substitution rate of 7.1×10^{-9} substitutions/site/year, which is the mean mutation rate in *A. thaliana* based on mutation accumulation experiments (Ossowski *et al.*, 2010). This substitution rate was used instead of providing calibration times for internal nodes, as fossil calibration times are not available for the divergence of *C. papaya* and *Vasconcellea* species. For each gene, the combined results of two runs with 10 million steps are presented.

The location of specific point mutations and changes in coding amino acids were documented using a self-written Perl script (Appendix S1).

Phylogenetic analysis

Multiple sequences were first aligned using the CLUSTALW program (Larkin *et al.*, 2007), and the alignments were exported to MEGA 4.0 (Tamura *et al.*, 2007) for neighbor-joining and maximum-parsimony analysis (both used 1000 replications for bootstrap). The alignments were also exported to MrBayes 3.1.2 (Huelsenbeck and Ronquist, 2001) for Bayesian analysis (General Time Reversible (GTR)+ invgamma model, 500 000 generations). The maximum-likelihood analysis was performed using PhyML 3.0 (Guindon and Gascuel, 2003) with a GTR model and the best of nearest neighbor interchanges (NNIs) tree topology. Homologs in *Arabidopsis thaliana* were used as an outgroup in all gene tree analyses.

Gene conversion analysis

We used GENECONV software version 1.81 (Sawyer, 1989, 1999) for gene conversion analysis. Both the global and pairwise comparisons were performed using 10 000 permutations. The fragments with higher significance than the Karlin–Altschul *P* value were recorded (Sawyer, 1989).

ACKNOWLEDGEMENTS

We thank Dr Matthew Hudson for revising the Perl scripts used in our analyses. We also thank the anonymous reviewers for their critical comments and generous suggestions. The research is supported by the National Science Foundation Plant Genome Research Program (award number 0553417). X.W. was supported by a fellowship from the Program in Physiological and Molecular Plant Biology of the University of Illinois.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article:

Figure S1. Positions of the five sex-linked BACs and structure of the six sex-linked genes.

Figure S2. Maximum-likelihood phylogenetic trees of sex-linked genes 1–6 and the autosomal *ANKYRIN* gene among *Carica* and *Vasconcellea* species.

Figure S3. Maximum-parsimony phylogenetic trees of sex-linked genes 1–6 and the autosomal *ANKYRIN* gene among *Carica* and *Vasconcellea* species.

Figure S4. Neighbor-joining phylogenetic trees of sex-linked genes 1–6 and the autosomal *ANKYRIN* gene among *Carica* and *Vasconcellea* species.

Figure S5. Single nucleotide polymorphisms (SNPs) in the coding sequences of genes 1–6 and the autosomal *ANKYRIN* gene.

Figure S6. Bayesian phylogenetic trees of sex-linked genes 1–6 and the autosomal *ANKYRIN* gene among *Carica* and *Vasconcellea* species after removing the sequences that showed gene conversion.

Table S1. Degenerate primers used to amplify the target genes.

Table S2. *Vasconcellea*-specific primers used to amplify the target genes and perform primer walking for sequencing the amplified fragments.

Table S3. Annotation of six sex-linked genes and the autosomal gene *ANKYRIN*.

Table S4. Estimates of synonymous and non-synonymous nucleotide divergence within and between species for gene 2.

Table S5. Estimates of synonymous and non-synonymous nucleotide divergence within and between species for gene 3.

Table S6. Estimates of synonymous and non-synonymous nucleotide divergence within and between species for gene 4.

Table S7. Estimates of synonymous and non-synonymous nucleotide divergence within and between species for gene 5.

Table S8. Estimates of synonymous and non-synonymous nucleotide divergence within and between species for gene 6.

Table S9. Fragments with possible gene conversion detected by GENECONV.

Appendix S1. Perl script for SNP analysis.

Please note: As a service to our authors and readers, this journal provides supporting information supplied by the authors. Such materials are peer-reviewed and may be re-organized for online delivery, but are not copy-edited or typeset. Technical support issues arising from supporting information (other than missing files) should be addressed to the authors.

REFERENCES

Aradhya, M.K., Manshardt, R.M., Zee, F. and Morden, C.W. (1999) A phylogenetic analysis of the genus *Carica* L. (Caricaceae) based on restriction fragment length variation in a cpDNA intergenic spacer region. *Genet. Resour. Crop Evol.* **46**, 579–586.

Badillo, V.M. (2000) *Carica* L. vs. Vasconcella St. Hil (Caricaceae): con la rehabilitacion de este ultimo. *Ernstia*, **10**, 74–79.

Charlesworth, B. (1991) The evolution of sex chromosomes. *Science*, **251**, 1030–1033.

Charlesworth, B. and Charlesworth, D. (1978) A model for the evolution of dioecy and gynodioecy. *Am. Nat.* **112**, 975–997.

Drummond, A.J. and Rambaut, A. (2007) BEAST: bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J. and Rambaut, A. (2006) Relaxed phyogenetics and dating with confidence. *PLoS Biol.* **4**, 699–710.

Filatov, D.A. and Charlesworth, D. (2002) Substitution rates in the X- and Y-linked genes of the plants, *Silene latifolia* and *S. dioica*. *Mol. Biol. Evol.* **19**, 888–907.

Fraser, J.A., Diezmann, S., Subaran, R.L., Allen, A., Lengeler, K.B., Dietrich, F.S. and Heitman, J. (2004) Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. *PLoS Biol.* **2**, e384.

Goddard, M.R., Godfray, H.C. and Burt, A. (2005) Sex increases the efficacy of natural selection in experimental yeast populations. *Nature*, **434**, 636–640.

Graves, J.A.M. and Shetty, S. (2001) Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. *J. Exp. Zool.* **290**, 449–462.

Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696–704.

Hall, T. (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser.* **41**, 95–98.

Huelsenbeck, J.P. and Ronquist, F. (2001) MrBayes: bayesian inferences of phylogeny. *Bioinformatics*, **17**, 754–755.

Kyndt, T., Romeijn-Peeters, E., Van Droogenbroeck, B., Romero-Motochi, J.P., Gheysen, G. and Goetghebeur, P. (2005) Species relationships in the

genus *Vasconcellea* (Caricaceae) based on molecular and morphological evidence. *Am. J. Bot.* **92**, 1033–1044.

Larkin, M.A., Blackshields, G., Brown, N.P. *et al.* (2007) ClustalW and ClustalX version 2. *Bioinformatics*, **23**, 2947–2948.

Liu, Z., Moore, P.H., Ma, H. *et al.* (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature*, **427**, 348–352.

Magdalita, P.M., Drew, R.A., Adkins, S.W. and Godwin, I.D. (1997) Morphological, molecular and cytological analyses of *Carica papaya* × *C. cauliflora* interspecific hybrids. *Theor. Appl. Genet.* **95**, 224–229.

Manshardt, R.M. and Wenslaff, T.F. (1989) Interspecific hybridization of papaya with other *Carica* species. *J. Am. Soc. Hort. Sci.* **114**, 689–694.

Ming, R. and Moore, P.H. (2007) Genomics of sex chromosomes. *Curr. Opin. Plant Biol.* **10**, 123–130.

Ming, R., Van Droogenbroeck, B., Moore, P.H., Zee, F.T., Kyndt, T., Scheldeman, X., Sekioka, T. and Gheysen, G. (2005) Molecular diversity of *Carica papaya* and related species. In *Plant Genome: Biodiversity and Evolution* (Sharma, A.K. and Sharma, A., eds). Enfield, NH: Science Publishers, pp. 230–254.

Ming, R., Yu, Q. and Moore, P.H. (2007a) Sex determination in papaya. *Semin. Cell Dev. Biol.* **18**, 401–408.

Ming, R., Wang, J.P., Moore, P.H. and Paterson, A.H. (2007b) Sex chromosomes in flowering plants. *Am. J. Bot.* **94**, 141–150.

Ming, R., Hou, S., Feng, Y. *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature*, **452**, 991–996.

Nei, M. and Gojobori, T. (1986) Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. *Mol. Biol. Evol.* **3**, 418–426.

Nicolas, M., Marais, G., Hykelova, V., Janousek, B., Laporte, V., Vyskot, B., Mouchiroud, D., Negrutiu, I., Charlesworth, D. and Monéger, F. (2005) A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol.* **3**, e4.

Ossowski, S., Schneeberger, K., Lucas-Lledó, J.I., Warthmann, N., Clark, R.M., Shaw, R.G., Weigel, D. and Lynch, M. (2010) The rate and molecular spectrum of spontaneous mutations in *Arabidopsis thaliana*. *Science*, **327**, 92–94.

Paland, S. and Lynch, M. (2006) Transitions to asexuality result in excess amino acid substitutions. *Science*, **311**, 990–992.

Peichel, C.L., Ross, J.A., Matson, C.K., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Mori, S., Schluter, D. and Kingsley, D.M. (2004) The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr. Biol.* **14**, 1416–1424.

Pike, A.C.W., Rellos, P., Niesen, F.H., Turnbull, A., Oliver, A.W., Parker, S.A., Turk, B.E., Pearl, L.H. and Knapp, S. (2008) Activation segment dimerization: a mechanism for kinase autophosphorylation of non-consensus sites. *EMBO J.* **27**, 704–714.

Rozas, J., Sánchez-DelBarrio, J.C., Messeguer, X. and Rozas, R. (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics*, **19**, 2496–2497.

Sawyer, S.A. (1989) Statistical tests for detecting gene conversion. *Mol. Biol. Evol.* **6**, 526–538.

Sawyer, S.A. (1999) GENECONV: a computer package for the statistical detection of gene conversion [WWW document]. URL http://www.math.wustl.edu/~sawyer.

Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J. *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837.

Spigler, R.B., Lewers, K.S., Main, D.S. and Ashman, T.L. (2008) Genetic mapping of sex determination in a wild strawberry, *Fragaria virginiana*, reveals earliest form of sex chromosome. *Heredity*, **101**, 507–517.

Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596–1599.

Van Droogenbroeck, B., Breyne, P., Goetghebeur, P., Romeijin-Peeters, E., Kyndt, T. and Gheysen, G. (2002) AFLP analysis of genetic relationship among papaya and its wild relative (Caricaeae) from Ecuador. *Theor. Appl. Genet.* **105**, 289–297.

Weismann, A. (1889) The significance of sexual reproduction in the theory of natural selection. In *Essays upon Heredity and Kindred Biological Problems* (Poulton, E.B., Schönland, S. and Shipley, A.E., eds). Oxford, UK: Clarendon Press, pp. 261–305.

Yin, T., DiFazio, S.P., Gunter, L.E. *et al.* (2008) Genome structure and emerging evidence of an incipient sex chromosome in *Populus*. *Genome Res.* **18**, 422–430.

Yu, Q., Hou, S., Hobza, R. *et al.* (2007) Chromosomal location and gene paucity of the male specific region on papaya Y chromosome. *Mol. Genet. Genomics*, **278**, 177–185.

Yu, Q., Hou, S., Feltus, F.A. *et al.* (2008a) Low X/Y divergence in four pairs of papaya sex-lined genes. *Plant J.* **53**, 124–132.

Yu, Q., Navajas-Perez, R., Tong, E., Robertson, J., Moore, P.H., Paterson, A.H. and Ming, R. (2008b) Recent origin of dioecious and gynodioecious Y chromosomes in papaya. *Trop Plant Biol.* **1**, 49–57.

Zhang, W., Wang, X., Yu, Q., Ming, R. and Jiang, J. (2008) DNA methylation and heterochromatinization in the male-specific region of the primitive Y chromosome of papaya. *Genome Res.* **18**, 1938–1943.

